DOI: 10.12731/2227-930X-2016-3-9-19

INTELLIGENT QUESTION-ANSWERING SYSTEM "MIVAR VIRTUAL CONSULTANT"

Adamova L.E., Protopopova D.A.

The paper deals with the process of designing question-answering system "Mivar Virtual Consultant" using specialized information-technology platform for understanding meaning of text in the natural Russian language. The system is capable of accumulating knowledge from texts in the natural Russian language and managing this knowledge. The methodology for training virtual consultant is described.

Keywords: *mivar; mivar networks; text meaning understanding; Mivar virtual consultant; artificial intelligence.*

1. Introduction

Currently question-answering systems are increasingly widespread. These are information systems capable of understanding questions and answering questions in a natural language [1–3]. Such systems provide the basis for designing virtual consultants (VC), which are applied to guide users through websites. VC can perform the following functions: communicating with users, guiding users through the website, virtual secretary, service desk or call-center specialist, etc.

The structure of VC can change depending on the approach used and the goals of the application. But in this systems it is possible to identify the following basic components:

- user interface for communication with the virtual consultant;
- the mechanism for processing user's questions and searching for answers;
- knowledge base of the virtual consultant.

The paper considers the process of designing the virtual consultant on the specialized platform Tel!Mi.

Virtual consultants refer to knowledge-based systems. The basis of such systems is structured semantic model of the subject domain, according to this model search for answer to the question is implemented. The source of information for models can be texts in a natural language. To convert text into such a model specialized systems are developed, that is, semantic parsers. There are different approaches to data storage and data accumulation in semantic models: semantic networks, conceptual dependencies, scenarios, frameworks, Resource Description Framework (RDF), etc. To solve tasks described above the platform Tel!Mi uses mivar-based approach [4–6], based on the following:

1) the technology of data representation on the basis of epistemological model VSO (the Russian abbreviation of the triple "Object-Property-Relation");
2) the technology of logical-and-computational data processing;
3) the use of previously accumulated models of subject domains (context).

The use of mivar-based technologies for designing virtual consultants allows us to carry out morphological, syntactical, and semantic text processing, as well as give answers to questions on the basis of knowledge [7–11]. It is known that there are methods of automatic question answering through search and retrieval of the answer phrase from huge collections of text information (IR technologies). These methods proved efficient for searching the necessary document relevant to the user's query in large text bases. This approach relies on enormous amount of information available as text on the Internet or specialized text collections. Having obtained the question from the user, the information retrieval techniques extract the answer in the form of a text fragment directly from these documents, guided by the text of the question.

Using this approach, firstly, the question is processed to determine the most likely answer type and queries for search engine are formulated. The search engine returns ranked documents divided into suit-

able passages. Finally, candidate answer strings are extracted from the passages and ranked.

However, users often want to get not the whole document, but a short answer to the question. To carry out accurate search for answers in the documents it is necessary to use methods that analyze text information flexibly and in more detail.

The majority of question-answering systems have a standard structure: they formulate a query from the user's question, implement query search to select parts of the documents that are most likely to contain the answer and then determine the most likely answers from these passages. Answers to short questions pose a problem to such systems operating on the word level (IR) and semantic-syntactic level (NLP), since it is difficult to find a few short answer segments. As a rule, IR methods lead to the use of huge statistics, whereas NLP methods result in entering a large number of rules into the knowledge base manually. To improve performance and adequacy of such systems it is necessary to increase the number of rules about the language and the world.

Mivar-based technologies allow us to implement natural Russian language understanding and give answers to questions on a higher semantic level. Firstly, mivar-based technologies allow us to process large texts, rather than separate sentences. Secondly, on preliminary stages of text processing mivar-based technologies allow us to use traditional syntactic parsers and other methods of traditional mathematical linguistics. Thirdly, mivar-based approach allows us to use several statistically trained traditional parsers simultaneously and compare the obtained results.

2. Mivar information-technology platform

The information-technology platform Tel!Mi has been developed on the basis of mivar technologies in the field of text meaning understanding in the natural Russian language. This platform allows us to store large amount of information, systematize and manage knowledge reasonably.

One of the goals of the system Tel!Mi is designing virtual consultants. Capabilities of VC depend on the set of knowledge that it has been taught. The set of knowledge of the consultant includes texts, glossaries, mivar networks with different type of connection and a list of predefined questions. Virtual consultant training in Tel!Mi is implemented using methodology for training a child, but the child is specific, he cannot touch an object, hold it in the palm, hear how a particular musical instrument sounds [10]. The training process is based on the methodology described in the next section.

Tel!Mi uses mivar-based technologies including the space VSO and mivar networks, which allows us to process texts in the natural Russian language. Tel!Mi allows us to develop the VSO model of a large enough text in short time. For example, the time of automated processing of a large text (Ojegov's Explanatory Dictionary) is:

0.05 (average processing time of one article) * 100 000 dictionary entries = 5000 seconds = 83 minutes.

It should be noted that Tel!Mi allows us to build a model of "Ojegov's Explanatory Dictionary" and other additional texts. Currently, total amount of knowledge base in the format of VSO model named "The picture of the world" accounts for 160 000 nodes and 600 000 arcs. Search for the answer to the user's question is executed on such an extremely large VSO model in less than a minute. In fact, the tasks of searching for a subgraph in the graph and finding the path between the nodes of the graph are solved.

Information-technology platform has allowed us to reduce text adaptation for virtual consultant training. Therefore, mivar virtual consultant is capable of operating texts in a living natural language.

3. The process of virtual consultant development using mivar-based technologies

The software suite of question answering system based on mivar technologies will be named "mivar virtual consultant" (MVC).

The process of developing the virtual consultant on the basis of the information technology platform Tel!Mi using mivar-based technologies can be divided into several stages:

1) Context analysis of the subject domain.

This is the major stage in the process of designing MVC, since it determines the specific subject domain of the virtual consultant. Depending on the selected subject domain, VC will have different amount of knowledge.

2) Developing and training the mivar knowledge base (KB) of the mivar virtual consultant. KB of the virtual consultant is built on the basis of concepts and networks. Concept is regarded as a word or a word combination that conveys the meaning of the term or terms. Concept network is connections between concepts in the picture of the world [10]. To form the concept network it is necessary to teach the virtual consultant's knowledge base concepts according to the analysis of the subject domain and interconnect these concepts into a single network, that is, the picture of the world. If the context is unambiguous in the selected subject domain, the process of developing concept network is considerably simplified.

If there is a lack of data obtained on the second stage, additional training should be done through concepts similar to the previous stage or through texts. To provide training through texts graphemic, morphological, syntactic and semantic analysis of input information should be carried out. Graphemic analysis is regarded as segmentation of the obtained text into paragraphs, each paragraph is divided into sentences, sentences are divided into word combinations and words. Then morphological analysis should be carried out in each sentence obtained as a result of graphemic analysis. The result is an array of word forms with a set of morphological characteristics (initial form, gender, number, etc.) within one sentence. In the process of morphological analysis Tel!Mi uses a morphological dictionary, which contains word forms with a set of morphological characteristic. Current-

ly, such a dictionary contains more than 3 million word forms. If in the process of morphological analysis the required word form is not found, the possibility of expanding morphological dictionary is provided. Moreover, automatic word form generation is provided taking into account rules of the Russian language (declension, conjugation). Then syntactic analysis is carried out, when using the sentence the graph is build that connect words in the sentence. The final stage of text analysis is semantic analysis, that is, defining the meaning of the entered information. Access to the concepts and concept networks available in the knowledge base of MVC is implemented. The steps of parsing the sentence "A boy found a beautiful flower under the tree" are represented in (**Fig. 1, 2, 3**).

3) Designing the dialog box interface.

The implementation of this stage depends on technical specification and type of the virtual consultant in particular (personalised, impersonalised, technical). If VC is personalised, the system Tel!Mi provides training in answering general questions and personal questions.

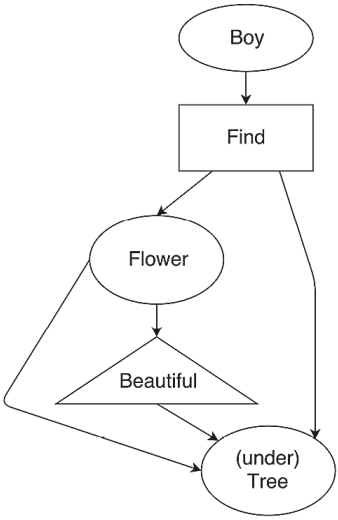

Fig. 1. Words

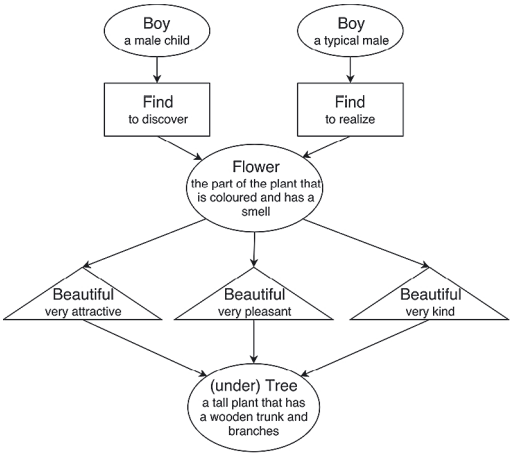

Fig. 2. Concepts

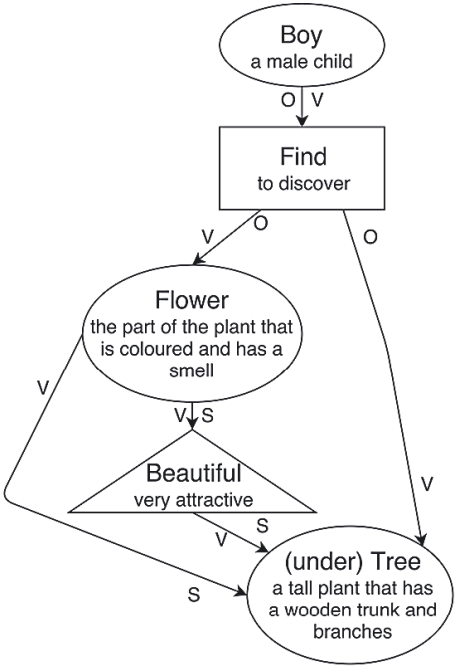

Fig. 3. VSO-model

4) Developing answer output algorithms.

Answer output algorithms depend on the type of user's query (a question to the consultant). The first stage involves the analysis of question type.

The following types of questions are identified:

– general questions, that is, questions that do not refer to the subject domain
– personal questions
– subject domain questions
– approval/ disapproval scenarios and etc.
– greeting/goodbye

After determining the question type, scenario manager, that is answer output algorithm, runs the appropriate scenario. Scenario is regarded as a sequence of steps to search for an answer to the question of a particular type. The result is the output of a certain answer in the consultant's dialogue box.

5) Testing.

The testing stage involves verifying correctness of data stored in the knowledge base of the MVC and debug of answer output algorithms. Using Tel!Mi allows us to design VC with required functions for entirely different subject domains.

4. the principles of methodology for training virtual consultant based on mvar technologies

MVC is necessary for human-computer communication, for example, answering questions on the websites, portals and analogous modern automatic systems. MVC development is based on using children training techniques on the basis of cause-effect rules (production rules) and thorough descriptions of the outside world.

MVC training is implemented constantly, continuously considering context involved in the training. Thus, context is a social environment of the virtual consultant. The context in the process of MVC

training is as important as in real human communication. MVC has its knowledge base, on the basis of which context understanding is implemented. Context understanding is based on the main principles and methods of remedial and rehabilitation training.

Virtual consultant training has its methodology developed on the basis of analysis of children training techniques and specific features of mivar-based technologies. This methodology involves several principles:

- teaching VC data in the form of dictionary entries;
- building mivar networks using texts;
- teaching VC groups of answers and questions;
- identifying mivar networks from dictionary entries with different types of relation (general –particular, the part –the whole, etc.)
- teaching VC concepts.

5. Conclusions

The application of mivar-based technologies to developing virtual consultants can boost operation of contact-centers and services aimed at mass market. Communication between the user and the virtual consultant is implemented on the semantics level. Virtual consultants designed using Tel!Mi are easily trained: they are capable of understanding written human speech, texts of books, adapted literary works. On the basis of taught knowledge MVC is capable of making decisions in different situations of communication with service's user.

References

1. Luger G.F. (2008), Artificial intelligence: structures and strategies for complex problem solving. Boston. Pearson education.
2. Burlak S. (2012), Origin of Language: Facts, Research, and Hypotheses. Astrel: CORPUS, Moscow.
3. Plangian V.A. (2012), Why are languages so different. AST-PRESS KNIGA, Moscow.

4. Varlamov O.O. (2002), Evolutionary databases and knowledge bases for adaptive synthesis of intelligent systems. Mivar-oriented information space. Radio and communications, Moscow, 288 p.
5. Varlamov O.O. (2002), "The fundamentals of multi-dimensional information developing (mivar) space of data and rule representation", *Information technologies*, No.5, pp. 42–47.
6. Varlamov O.O. (2015), "Mivar role and place in computer sciences, artificial intelligence systems and informatics", *Radio industry*, No. 3, pp. 10–27.
7. Varlamov O.O., Adamova L.E., Eliseev D.V., Mayboroda Yu.I., Antonov P.D., Sergushin G.S., Chibirova M.O. (2013), Comprehensive modelling of the processes of text meaning understanding, speech and image understanding by computers on the basis of mivar technologies. *Artificial intelligence*, No. 4, pp. 15–27.
8. Varlamov O.O., Adamova L.E., Peterson A.O., Protopopova D.A., Skakunova E.A. (2014), Research of approaches and major problems of natural Russian language understanding. *Automation and control in technical systems*, No. 2, available at: URL: auts.esrae.ru/10-196.
9. Varlamov O.O., Adamova L.E., Peterson A.O., Protopopova D.A., Skakunova E.A. (2014), Mivar text understanding: developing methodology for teaching virtual personality domain knowledge on the basis of concept network design. *Automation and control in technical systems*, No.3, available at: URL: auts.esrae.ru/11-210.
10. Varlamov O.O. (2015), "Formalization of the term "text meaning understanding" on the basis of mivar technologies and the concept "Thing-Property-Relation", *Radio industry*, No. 3, pp. 144–159.
11. Adamova L.E., Belousova A.I., Protopopova D.A., Eliseev D.V., Peterson A.O. (2015), "An approach to creating intelligent question-answering system "Mivar Virtual Consultant", *Radio industry*, No. 3, pp. 160–171.
12. Hermjakob U., Hovy E.H., Lin C.Y. (2000), Knowledge-based question answering, *Proceedings of the Sixth World Multiconference on Systems, Cybernetics, and Informatics (SCI-2002).*

DATA ABOUT THE AUTHORS

Adamova Larisa Evgen'evna, Deputy Director General, PhD in Psychology
LLC "Mivar"
72, Oktyabrskaya Str., Moscow, 127521, Russian Federation
l.adamova@mivar.ru

Protopopova Dar'ya Aleksandrovna, Student, Department of «Automated Control Systems»
State Technical University – MADI
64, Leningradsky prospekt, Moscow, 125319, Russian Federation
d.protopopova@mivar.ru